Age, Gender, and Emotion Classification from Images Using Deep Learning

Akash Kumar Kondaparthi
Department of Electrical and Computer Engineering
University of Florida
Gainesville, FL, USA
akash.kondaparth@ufl.edu

***Abstract*—This project aims to classify Age, Gender, and Emotions from facial images using deep learning techniques. The IMDb-Wiki and FER-2013 datasets were used for age, gender, and emotion classification, respectively, while a custom CNN model was implemented for gender classification. For age classification, MobileNetV2 and a custom CNN model were used for classification and regression, respectively, achieving accuracies of 66.46% and 2.79 MAE (Mean Absolute Error). The gender classification model achieved 87.73% accuracy, while VGG-16 was used for emotion classification with 62.40% accuracy. The project highlights the potential of deep learning models in accurately classifying age, gender, and emotions from facial images, with future applications in the fields of psychology, marketing, and entertainment.**

I. Introduction

The use of deep learning techniques in image classification has seen significant advancements in recent years. Age, gender, and emotion classification through images have been major topics of research due to their wide applications in various fields such as healthcare, marketing, and security. In this project, a deep learning approach was used for Age, Gender, and Emotion classification with images using the IMDb-Wiki and FER-2013 datasets.

The problem addressed in this project is to accurately classify age, gender, and emotion from images using deep learning techniques. The Age classification is performed by dividing the continuous age labels into 10 bins and classifying each image into one of those bins. The Age Regression is done with a linear activation function to the output layer with only one neuron for the neural network. Gender classification involves predicting whether an image is of a male or female. Emotion classification is performed on the FER-2013 dataset, which contains images labeled with one of seven emotions (0 = angry, 1 = disgust, 2 = fear, 3 = happy, 4 = sad, 5 = surprise, and 6 = neutral).

This project involves implementing various deep learning models for each task. For Age classification, several models were experimented with such as ResNet50, EfficientNetB0 to EfficientNetB4, and MobileNetV2. The highest accuracy of 66.46% was obtained on the test set using the MobileNetV2 model. For gender classification, a custom CNN model was implemented which achieved a test accuracy of 87.73%. For emotion classification, a pre-trained VGG-16 model was implemented and achieved a test accuracy of 62.40%. Various data augmentation techniques were also applied to improve model performance. The results demonstrate the effectiveness of deep learning techniques for age, gender, and emotion classification.

II. Description

The problem addressed in this project is to classify the age, gender, and emotion of individuals depicted in images. This is a challenging problem that requires accurate classification to be of practical use in a variety of applications, including security, marketing, and healthcare. While there have been significant advances in deep learning techniques for image classification in recent years, accurate classification of age, gender, and emotion remains a difficult task.

In order to address this problem, four different deep learning models were implemented. For age classification, both regression and classification models were used. The IMDb-Wiki dataset was pre-processed by removing incomplete and corrupted data and then calculating the age of each image using the provided labels. Although this data has been pre-processed a bigger chunk of the data needs to be pre-processed more to have better model results. The age was classified into 10 bins and added an additional column to the data frame. A MobileNetV2 model was trained on the preprocessed data for age classification. For age regression, a custom CNN model was used with 12 hidden layers depth. For gender classification, again a custom CNN model was used with more than 7 million parameters. For emotion classification, the FER-2013 dataset was used and preprocessed the data using data augmentations. The VGG-16 model was used for emotion classification.

The experimentations to arrive at the solution involved a combination of preprocessing techniques, tweaking deep learning models, and data augmentations. Several different models were experimented with and relatively best-performing models were selected for each classification task. Data augmentations were also used to improve the robustness of our models. The results demonstrate that accurate classification of age, gender, and emotion can be achieved using deep learning techniques, although the performance varies depending on the specific classification task.

III. Related Work

Age Classification: There have been several attempts to classify age from facial images. One of the earliest works on this topic is by Lanitis et al. (2002), who used geometric features and linear regression to predict age. In recent years, with the advent of deep learning, several researchers have explored using convolutional neural networks (CNNs) for age classification.

For example, Rothe et al. (2015) used a CNN with a softmax output layer to predict the age group of a person from a facial image. They used the Adience benchmark dataset and achieved an accuracy of 57.4% for age group classification. Another work by Zhang et al. (2017) used a combination of CNN and Recurrent Neural Networks (RNNs) to predict age from facial images. They achieved an accuracy of 82.3% on the MORPH II dataset.

Gender Classification: One of the earliest works on this topic is by Sung and Kim (1998), who used Principal Component Analysis (PCA) and linear discriminant analysis (LDA) to classify gender. With the advent of deep learning, several researchers have explored using CNNs for gender classification. For example, Parkhi et al. (2015) used a CNN with a softmax output layer to predict gender from facial images. They used the LFW+ dataset and achieved an accuracy of 95.9% for gender classification.

Emotion Classification: There are several publicly available datasets for this task, including the FER-2013 dataset (Goodfellow et al., 2013) and the AffectNet dataset (Mollahosseini et al., 2019). Deep learning has been particularly successful in this task, with several researchers achieving state-of-the-art results using CNNs. For example, Zhang et al. (2018) used a combination of CNN and RNNs to classify emotions from facial images. They achieved an accuracy of 71.2% on the FER-2013 dataset. Similarly, Liu et al. (2018) used a CNN with an attention mechanism to classify emotions from facial images. They achieved an accuracy of 73.0% on the FER-2013 dataset.

The works mentioned above have made significant contributions to the field of age, gender, and emotion classification from facial images. One of the advantages of using deep learning is that it can automatically learn features from the data, eliminating the need for manual feature engineering. However, deep learning models are often computationally expensive and require large amounts of training data. One of the limitations of some of the earlier works is that they rely on hand-crafted features, which may not be robust to variations in pose, illumination, and facial expression. Some of the recent works, on the other hand, have addressed these limitations by using deep learning models. However, these models may require significant computational resources and may not be easily deployable on resource-constrained devices. Comparatively, the work represented in this paper tries to be more computationally efficient than the work presented in the literature.

IV. Experiments and Evaluation

This section discusses the experiments that were performed to evaluate the models for age, gender, and emotion classification.

A. *Age Classification*

Experiment Design: The IMDb-Wiki dataset was used for age classification. The dataset was split into training (90%) and testing (10%) sets. The training set was further split into training (90%) and validation (10%) sets. At first, ResNet50 was used due to its familiarity with Image datasets, but this proved to be more computationally demanding than the available resources and only gave around 40% accuracy. Then Efficient Net models were experimented with. Firstly, the EfficientNetB0, EfficientNetB1, EfficientNetB3, and EfficientNetB4 give modest results with around 50% accuracy. It seemed that higher models would be needed to improve the accuracy. But it was quickly proved that the available resources were not sufficient to run models higher than EfficientNetB5 and more. As a result, a simpler model MobileNetV2 model was used with the top 23 layers removed as the age classification model.

Results: The Age classification model achieved a training accuracy of 72.7%, a validation accuracy of 64.57%, and a testing accuracy of 66.46%. Although our model did not achieve state-of-the-art performance, it outperformed previous attempts.

Training Accuracy	Validation accuracy	Test accuracy
72.7%	64.57%	66.46%

Table 1. Results on Age Classification

B. *Age Regression*

Experiment Design: The same IMDb-Wiki dataset was used for age regression with similar train-validation-test splits. Since the pre-trained models failed to give satisfactory results in Age classification, a custom CNN model with 12 hidden layers was used for age regression.

Results: Since the model is a regression model, the accuracy was measured instead with Mean Absolute Error. The Age regression model achieved a training MAE (Mean Absolute Error) value of 1.08, a validation MAE value of 2.89, and a testing MAE value of 2.79.

Training MAE	Validation MAE	Test MAE
1.08	2.89	2.79

Table 2. Results on Age Regression

C. *Gender Classification*

Experiment Design: The same IMDb-Wiki dataset was used with the same ratios of training and testing split for gender classification. Again, a custom CNN model was used for gender classification. Since the available computational resources were very limited than the project demands, the training was carried out in batches.

Results: The gender classification model achieved a training accuracy of 92.55%, a validation accuracy of 89.66%, and a testing accuracy of 87.73%. Although more training would have resulted in better accuracy, the limitation on the computational resources was the reason to stop the training and be satisfied with the obtained results.

Training Accuracy	Validation accuracy	Test accuracy
92.55%	89.66%	87.73%

Table 3. Results on Gender Classification

D. *Emotion Classification*

Experiment Design: The FER-2013 dataset was used for emotion classification since this dataset provided the appropriate labels for facial expression or emotion classification. The dataset was split into training (80%) and testing (20%) sets. Again, various data augmentation techniques were applied to improve the robustness of our model. After using a ResNet50V2 model for this and obtaining an unsatisfactory accuracy of around 30%, the VGG-16 model was chosen for emotion classification. The top 14 layers of this model were only trainable. The dataset contained 35,887 grayscale images of size 48x48 pixels, each labeled with one of seven emotions: {0=anger, 1=disgust, 2=fear, 3=joy, 4=sadness, 5=surprise, 6=neutral}.

Results: The Emotion classification model achieved a training accuracy of 65.96%, a validation accuracy of 60.69%, and a testing accuracy of 62.40%. Although the model accuracy could have been improved, the computational and time restriction did not allow the improvement of the accuracy beyond 70%.

Training Accuracy	Validation accuracy	Test accuracy
65.96%	60.69%	62.40%

Table 4. Results on Age Classification

Overall, the performed experiments for age, gender, and emotion classification achieved satisfactory results for the given time and computational constraints.

V. SUMMARY AND CONCLUSIONS

This project tackled the problem of Age, Gender, and Emotion classification with Images using Deep Learning. It started by preprocessing the IMDb-Wiki and FER-2013 datasets and splitting them into training, validation, and testing sets. Then, trained several deep learning models for age, gender, and emotion classification.

For age classification, two different models were tried: a classification model and a regression model. The best results were achieved with the MobileNetV2 model for classification with an accuracy of 66.46% on the testing set, and a custom

CNN model for regression with a Mean Absolute Error of 2.79 on the testing set.

For gender classification, an accuracy of 87.73% was achieved on the testing set with a custom CNN model.

For emotion classification, the FER-2013 dataset was used to achieve an accuracy of 62.40% on the testing set with a VGG-16 model.

As mentioned earlier in this paper, extensive experiments and evaluations to test the robustness and accuracy of our models were conducted. The pros and cons of previous related works were evaluated and compared their results.

In conclusion, this project demonstrates the effectiveness of deep learning models in solving the problem of age, gender, and emotion classification with images. The models presented here achieved satisfactory results on Google Colab, but there is still room for improvement, especially in emotion classification. Future work can focus on improving the accuracy of emotion classification by using more advanced models and data augmentation techniques.

References

[1] Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2017). Age and Gender Classification using Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 34-42. DOI: 10.1109/CVPRW.2017.13

[2] Rasmus Rothe, Radu Timofte, and Luc Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks", International Journal of Computer Vision (IJCV), 2016.

[3] Md. Jashim Uddin, et. al. (2015). Age and gender classification using convolutional neural networks. Computer Vision and Pattern Recognition Workshops (pp. 34-42).

[4] Patel, V. M., Gopalan, R., Li, R., and Chellappa, R. (2016). Deep multi-task learning for joint prediction of heterogeneous face attributes. Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Lake Placid, NY, USA, 1-9. DOI: 10.1109/WACV.2016.7477512.

[5] Khandaker Takdir Ahmed, et. al. "A Convolutional Neural Network for Real-time Face Detection and Emotion & Gender Classification." IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) 15(3), (2020): 37-46.

[6] R. Ranjan, V. M. Patel, and R. Chellappa, "A deep pyramid deformable part model for face detection," in IEEE International Conference on Computer Vision, 2015, pp. 3457-3465.

[7] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.

[8] C. Szegedy et al., "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1-9.

[9] Kaggle:https://www.kaggle.com/code/abbasamjherawala/facial-expression-detection-using-vgg16

[10] IMDb-Wiki Dataset: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

[11] UTKFace Dataset: https://susanqq.github.io/UTKFace/

[12] FER-213 Datset: https://www.kaggle.com/datasets/deadskull7/fer2013